

Comparative Genomic Analysis Reveals a Distant Liver Enhancer Upstream of the *COUP-TFII* gene

Nadine Baroukh^{*&}, Nadav Ahituv^{*°&}, Jessie Chang^{*}, Malak Shoukry^{*}, Veena Afzal^{*},
Edward M. Rubin^{*°} and Len A. Pennacchio^{*°}

[&]Both authors contributed equally to this work.

^{*}Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley,
California 94720 USA. [°]U.S. Department of Energy Joint Genome Institute, Walnut
Creek, California 94598 USA.

To whom correspondence should be addressed:

Len A. Pennacchio, Genomics Division, One Cyclotron Road, MS 84-171, Lawrence
Berkeley National Laboratory, Berkeley, CA 94720. Email: LAPennacchio@lbl.gov,
Phone: (510) 486-7498, Fax: (510) 486-4229

Running title: *A liver enhancer near COUP-TFII*

Abstract

COUP-TFII is a central nuclear hormone receptor that tightly regulates the expression of numerous target lipid metabolism genes in vertebrates. However, it remains unclear how *COUP-TFII* itself is transcriptionally controlled since studies with its promoter and upstream region fail to recapitulate the genes' liver expression. In an attempt to identify liver enhancers in the vicinity of *COUP-TFII*, we employed a comparative genomic approach. Initial comparisons between humans and mice of the 3,470kb gene poor region surrounding *COUP-TFII* revealed 2,023 conserved non-coding elements. To prioritize a subset of these elements for functional studies, we performed further genomic comparisons with the orthologous pufferfish (*Fugu rubripes*) locus and uncovered two anciently conserved non-coding sequences (CNS) upstream of *COUP-TFII* (CNS-62kb and CNS-66kb). Testing these two elements using reporter constructs in liver (HepG2) cells revealed that CNS-66kb, but not CNS-62kb, yielded robust *in vitro* enhancer activity. In addition, an *in vivo* reporter assay using naked DNA transfer with CNS-66kb linked to luciferase displayed strong reproducible liver expression in adult mice, further supporting its role as a liver enhancer. Together, these studies further support the utility of comparative genomics to uncover gene regulatory sequences based on evolutionary conservation and provide the substrates to better understand the regulation and expression of *COUP-TFII*.

Introduction

Transcriptional control provides a central homeostatic mechanism for regulating lipid metabolism. Orphan steroid hormone receptors play a key role in this process and represent an important set of ligand-activated transcription factors. One family of this class include the Chicken Ovalbumin Upstream Promoter-Transcription Factors (COUP-TFs) (Sagami et al. 1986; Wang et al. 1987) which were first identified as homodimers that bind to a direct GGTC A repeat regulatory element in the chicken ovalbumin promoter (Sagami, Tsai et al. 1986). Two distinct COUP-TFs have been identified in mammals, *COUP-TFI/EAR3/NR2F1* (Miyajima et al. 1988; Wang et al. 1989) and *COUP-TFII/ARPI/NR2F2* (Ldias and Karathanasis 1991; Wang et al. 1991), as well as numerous homologues in species ranging from flies to humans. These two *COUP-TF* paralogues are expressed in many tissues during various developmental stages (Miyajima, Kadowaki et al. 1988; Ldias and Karathanasis 1991) and are involved in the regulation of numerous important biological processes, including neurogenesis, organogenesis, cell fate determination and metabolic homeostasis (Wang, SY et al. 1989; Ritchie et al. 1990; Wang, Ing et al. 1991; Qiu et al. 1994).

Specifically, *COUP-TFII* is expressed in restricted regions of the central nervous system during embryonic development, as well as mesenchymal tissue of many organs and in the developing eyes (Qiu et al. 1994; Qiu et al. 1995). However, the *cis*-regulatory sequences responsible for this regulation remain partially defined. Using a 4.2kb mouse DNA fragment of the *COUP-TFII* promoter plus the upstream interval, Zhang et al.

(Zhang et al. 2002) tested the regulatory properties of this sequence in transgenic mouse embryos. At embryonic day 9, expression was found in the notochord, the ventral neural tube, and the gut endoderm, while at embryonic day 15.5 *COUP-TFII* expression was found in numerous tissues derived from endoderm, including lung, pancreas, stomach, intestine, bladder, esophagus and thyroid. However, no expression was found in liver despite *COUP-TFII*'s known expression in this tissue at this embryonic developmental stage. These data suggested the existence of a distant *COUP-TFII* liver enhancer which we set out to identify in this study. Based on the availability of several vertebrate genome sequences and the success of comparative genomics to identify functional sequences (Pennacchio and Rubin 2001; Nobrega and Pennacchio 2004; Zheng et al. 2004), we used an evolutionary conservation filter to identify conserved non-coding sequences in vicinity of *COUP-TFII* and subsequently tested these sequences for liver enhancer properties.

Material & Methods

Comparative Genomics and Human *COUP-TFII* plasmid constructs

Human, mouse, and fugu sequences were obtained from the UCSC genome browser (<http://genome.ucsc.edu/>) and genomic sequence comparisons were performed using VISTA (Mayor et al. 2000). We cloned the conserved noncoding human sequences (CNS-62kb and CNS-66kb) upstream of the pGL3-Promoter vector (Promega) using the

following primers: CNS-66kb-F: CCTCTGGGCACACGGACTCT; CNS-66kb-R: CTAATAAAGTCCAAGGG TC; and CNS-62kb-F:ATGTTTGTGGCGTTTCTTCC; CNS-62kb-R:TA ACTCAAGCCTCAGAAAGC.

Reporter gene transfection studies: Luciferase Assays

Human HepG2 hepatoma cells (ATCC HB8065) were grown in DMEM (GibcoBRL) supplemented with 10% FBS (Hyclone), 2mM glutamine, 0.1mM non-essential amino acids, and 1mM sodium pyruvate. Cells were grown in 6 well plates and transfected using Fugene (Roche Molecular Biochemicals) following the manufacturer's protocol. Transfections were carried out in triplicates. Briefly, 1 µg of each assayed plasmid and 0.25 µg of pSEAP2-control plasmid (BD Biosciences) were mixed with 6 µl Fugene and added to each well. Following 48 hours of incubation cells were assayed for transfection efficiency using SEAP with the Bright-Glo Luciferase Assay System (Promega) and then harvested and lysed using the Glo Lysis Buffer (Promega). Luciferase was assayed using Bright-Glo Luciferase Assay System (Promega) and results were calculated as fold difference from the pGL3-Promoter parent vector.

Tail vein plasmid DNA transfer assays

Each separate tail vein injection experiment was performed as described by Herweijer and Wolff (Herweijer and Wolff 2003) on 3 CD1 mice using and following the TransIT[®] *In Vivo* Gene Delivery System Protocol (Mirus Corporation). Briefly, 10 µg of each

clone, along with 2 µg of pCMVβ (BD Biosciences) to correct for delivery efficiency, were injected into each mouse. Animals were sacrificed 24 hours later, livers extracted, measured to correct for size, homogenized, and centrifuged for 10 minutes at 4,000 rpm. Activity of luciferase and LacZ was measured using the Luciferase Assay System (Promega) and the Galacto-Light Plus (Applied Biosystems) respectively, and quantified using the Lumat LB 9501 (Berthold).

Results

To identify potential *COUP-TFII* regulatory sequences, we performed comparative sequence analysis and scanned for evolutionary conserved noncoding sequences. Initially, we compared the human *COUP-TFII* interval with orthologous mouse sequence and identified 2,023 conserved noncoding sequences spread throughout two large gene poor regions flanking the gene, that encompass only two genes, *COUP-TFII* and *MGC44294* (Figure 1A). These include 1,850kb and 1,620kb regions upstream and downstream of the gene, respectively (Figure 1A).

To further prioritize these human-mouse conserved noncoding sequences for functional studies, we added a requirement for the sequence to be conserved in a more distant species. Using the available *Fugu rubripes* *COUP-TFII* genomic sequence the number of conserved noncoding sequences in this region was reduced to three human/mouse/fugu elements showing >70% identity over >90bp (Figure 1B). All three elements were

located upstream of the transcription start site, with one corresponding to the known proximal promoter of the gene. Thus, two novel intervals (CNS-66kb and CNS-62kb) upstream of the gene became candidates for potential gene regulatory activity based on their evolutionary conservation. Comparison between human/fugu revealed CNS-66kb was 193bp with 84% identity and CNS-62kb was 94bp with 76% identity.

As an initial test for potential liver enhancers, we examined both elements using a luciferase reporter vector transfection assays in liver cells (HepG2). While minimal activity was found for CNS-62kb, CNS-66kb revealed 6-fold higher luciferase expression compared to the minimal promoter luciferase plasmid (Figure 2). These data suggested that human CNS-66kb could function as a liver enhancer, at least based on this *in vitro* assay.

To assess the ability of CNS-66kb to act as a liver enhancer *in vivo*, we explored naked DNA transfer of luciferase reporter vector constructs through tail vein injection of adult mice (Herweijer and Wolff 2003). This method has been shown to deliver foreign DNA to the liver as a result of the use of a large volume of injection material (10% of the mouse weight) in a very short period of time (6-8 seconds). In these experiments CNS-66kb yielded robust reporter vector expression with levels ~40-fold higher than the basal promoter construct and ~8-fold higher than the well-defined ApoE liver enhancer control (Figure 3).

Discussion

The identification of gene regulatory elements embedded in >95% of the human genome which is non-coding remains a significant challenge. This is exemplified in the case of *COUP-TFII*, where several of the key gene regulatory elements remain undefined. The previous finding that the immediate flanking upstream sequence (-4.2 kb) of *COUP-TFII* failed to drive expression in liver tissue (Zhang, Bennoun et al. 2002) provided support that other distant *cis*-regulatory elements are involved in its expression. In this study we explored the flanking *COUP-TFII* interval and were challenged with a large genomic search space composed of only two genes in a region which covers ~3500 kb, where gene regulatory sequences could reside.

While traditional deletion constructs of large genomic transgenes have identified a handful of distant regulatory sequences, the availability of several vertebrate genomes has enabled comparative genomic approaches to increasingly facilitate the identification of distant gene regulatory sequences (Boffelli et al. 2004). In this study based on human/mouse/fugu evolutionary conservation, we were able to identify three conserved non-coding sequences within the ~3500 kb region flanking *COUP-TFII*. While one overlapped with the well-defined proximal promoter (Zhang, Bennoun et al. 2002), two were located approximately 62 and 66 kb upstream of the gene and had not been previously described. Using two relatively rapid reporter vector construct assays we were able to identify CNS-66kb as a putative liver enhancer. Interestingly, CNS-66kb showed even higher *in vivo* liver expression levels of a reporter gene when compared to

the well-characterized and robust *APOE* liver enhancer (Shachter et al. 1993; Simonet et al. 1993). These observations warrant further investigation and suggest that this sequence may be useful to drive robust liver expression in future applications such as hepatic gene therapy or conditional gene knockouts.

Acknowledgements

This work was funded in part by the NIH-NHLBI Programs for Genomic Application Grant HL66681 (EMR) and HL071954A (LAP) through the U.S. Department of Energy under contract no. DE-AC03-76SF00098, and an American Heart Association Postdoctoral Fellowship (NB).

References

- Boffelli D, Nobrega M A, Rubin E M (2004) Comparative genomics at the vertebrate extremes. *Nat Rev Genet* 5, 456-465
- Herweijer H, Wolff J A (2003) Progress and prospects: naked DNA gene transfer and therapy. *Gene Ther* 10, 453-458
- Ladias J A, Karathanasis S K (1991) Regulation of the apolipoprotein AI gene by ARP-1, a novel member of the steroid receptor superfamily. *Science* 251, 561-565
- Mayor C, Brudno M, Schwartz J R, Poliakov A, Rubin E M, et al. (2000) VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16, 1046-1047
- Miyajima N, Kadowaki Y, Fukushige S, Shimizu S, Semba K, et al. (1988) Identification of two novel members of erbA superfamily by molecular cloning: the gene products of the two are highly related to each other. *Nucleic Acids Res* 16, 11057-11074
- Nobrega M A, Pennacchio L A (2004) Comparative genomic analysis as a tool for biological discovery. *J Physiol* 554, 31-39
- Pennacchio L A, Rubin E M (2001) Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* 2, 100-109
- Qiu Y, Cooney A J, Kuratani S, DeMayo F J, Tsai S Y, et al. (1994) Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-transcription factors in the developing mouse central nervous system: evidence for a role in segmental patterning of the diencephalon. *Proc Natl Acad Sci U S A* 91, 4451-4455

- Qiu Y, Krishnan V, Zeng Z, Gilbert D J, Copeland N G, et al. (1995) Isolation, characterization, and chromosomal localization of mouse and human COUP-TF I and II genes. *Genomics* 29, 240-246
- Qiu Y, Tsai S Y, Tsai M J (1994) COUP-TF: An orphan member of the steroid/thyroid hormone receptor superfamily. *Trends Endocrinol Metab* 5, 234-239
- Ritchie H H, Wang L H, Tsai S, O'Malley B W, Tsai M J (1990) COUP-TF gene: a structure unique for the steroid/thyroid receptor superfamily. *Nucleic Acids Res* 18, 6857-6862
- Sagami I, Tsai S Y, Wang H, Tsai M J, O'Malley B W (1986) Identification of two factors required for transcription of the ovalbumin gene. *Mol Cell Biol* 6, 4259-4267
- Shachter N S, Zhu Y, Walsh A, Breslow J L, Smith J D (1993) Localization of a liver-specific enhancer in the apolipoprotein E/C-I/C-II gene locus. *J Lipid Res* 34, 1699-1707
- Simonet W S, Bucay N, Lauer S J, Taylor J M (1993) A far-downstream hepatocyte-specific control region directs expression of the linked human apolipoprotein E and C-I genes in transgenic mice. *J Biol Chem* 268, 8221-8229
- Wang L, SY T, RG C, WG B, MJ T, et al. (1989) COUP transcription factor is a member of the steroid receptor superfamily. *Nature* 340, 163-166
- Wang L H, Ing N H, Tsai S Y, O'Malley B W, Tsai M J (1991) The COUP-TFs compose a family of functionally related transcription factors. *Gene Expr* 1, 207-216
- Wang L H, Tsai S Y, Sagami I, Tsai M J, O'Malley B W (1987) Purification and characterization of chicken ovalbumin upstream promoter transcription factor from HeLa cells. *J Biol Chem* 262, 16080-16086

- Zhang P, Bennoun M, Gogard C, Bossard P, Leclerc I, et al. (2002) Expression of COUP-TFII in metabolic tissues during development. *Mech Dev* 119, 109-114
- Zheng P, Pennacchio L A, Le Goff W, Rubin E M, Smith J D (2004) Identification of a novel enhancer of brain expression near the apoE gene cluster by comparative genomics. *Biochim Biophys Acta* 1676, 41-50

Figure Legends

Figure 1: The *COUP-TFII* genomic locus. A) A schematic representation of the human *COUP-TFII* region. B) A VISTA plot showing the human-mouse and human-fugu conservation in the 5' region adjacent to *COUP-TFII*. Arrows indicate the location of CNS-66kb and CNS-62kb.

Figure 2: *COUP-TFII* *in vitro* enhancer assay. A bar plot showing the luciferase transcriptional activation levels, adjusted to the background levels of the transfected control construct pGL3-Promoter vector. * <0.05 and ** <0.001 p-value (independent t-test).

Figure 3: *COUP-TFII* *in vivo* enhancer assay. A bar plot showing the luciferase transcriptional activation levels, adjusted to the background levels of the injected control construct pGL3-Promoter vector. ** <0.001 p-value (independent t-test).